

# GENERAL FAMILIES OF CHAIN RATIO TYPE ESTIMATORS OF THE POPULATION MEAN WITH KNOWN COEFFICIENT OF VARIATION OF THE SECOND AUXILIARY VARIABLE IN TWO PHASE SAMPLING

HOUSILA P. SINGH<sup>1</sup>, SARJINDER SINGH<sup>2</sup> AND JONG-MIN KIM<sup>3</sup>

## ABSTRACT

In this paper we have suggested a family of chain estimators of the population mean  $\bar{Y}$  of a study variate  $y$  using two auxiliary variates in two phase (double) sampling assuming that the coefficient of variation of the second auxiliary variable is known. It is well known that chain estimators are traditionally formulated when the population mean  $\bar{X}_1$  of one of the two auxiliary variables, say  $x_1$ , is not known but the population mean  $\bar{X}_2$  of the other auxiliary variate  $x_2$  is available and  $x_1$  has higher degree of positive correlation with the study variate  $y$  than  $x_2$  has with  $y$ ,  $x_2$  being closely related to  $x_1$ . Here the classes are constructed when the population mean  $\bar{X}_1$  of  $x_1$  is not known and the coefficient of variation  $C_{x_2}$  of  $x_2$  is known instead of population mean  $\bar{X}_2$ . Asymptotic expressions for the bias and mean square error (MSE) of the suggested family have been obtained. An asymptotic optimum estimator (AOE) is also identified with its MSE formula. The optimum sample sizes of the preliminary and final samples have been derived under a linear cost function. An empirical study has been carried out to show the superiority of the constructed estimator over others.

*AMS 2000 subject classifications.* Primary 65D05; Secondary 62J10.

*Keywords.* Study variate, auxiliary variates, population mean, coefficient of variation, bias, mean square error, double sampling.

---

Received November 2005; accepted September 2006.

<sup>1</sup>School of Studies in Statistics, Vikram University, Ujjain 456010, India (e-mail: hsingh@winedt.com)

<sup>2</sup>Department of Statistics and Computer Networking, St. Cloud State University, St. Cloud, MN 56301, U.S.A. (e-mail: sarjinder@yahoo.com)

<sup>3</sup>Corresponding author. Statistics, Division of Science and Mathematics University of Minnesota, Morris, MN 56267, U.S.A. (e-mail: jongmink@morris.umn.edu)

## 1. INTRODUCTION

In many sample surveys, the information on single (or more) auxiliary variable(s) correlated with the study variable is used for increasing the precision of estimators. A number of sampling strategies utilize the advance information about an auxiliary variable. When such information is lacking, it is sometimes relatively cheap to take a large preliminary sample in which auxiliary variable alone is measured. The aim of this sample is to obtain a good estimate of the population mean or total of the auxiliary variable or its frequency distribution (Prabhu-Ajgaonkar, 1975). This technique is known as double sampling or two-phase sampling. Neyman (1938) was the first to formulate it in connection with collecting information on all strata sizes in a stratified sampling. The use of two-phase sampling is necessary if the value of auxiliary variable is obtained on performing a nondestructive experiment whereas to obtain the value of the study variable of a unit a destructive experiment has to be performed, see Unnikrishnan and Kunte (1995). Sukhatme (1962) mentioned that in a survey to estimate the production of lime crop based on orchards as sampling units, a comparatively larger sample is taken to determine the acreage under the crop while the yield rate is determined from only a sub sample of the orchards selected for determining acreage. In addition to these, many other situations may be cited where use of double sampling is inevitable. Let  $U = (U_1, U_2, \dots, U_N)$  be a finite population of  $N$  units. Let  $y$  and  $x_1$  denote the study variate and auxiliary variate taking the values  $y_i$  and  $x_{1i}$  respectively on the unit  $U_i$  ( $i = 1, 2, \dots, N$ ). The problem of estimating the population mean  $\bar{Y}$  of  $y$  when the population mean  $\bar{X}_1$  of  $x_1$  is known, has been dealt with a great length in the literature. However, in certain practical situations when no information is available on the population mean  $\bar{X}_1$  of  $x_1$  before the start of the survey, we seek to estimate  $\bar{Y}$  from a sample  $s$  obtained through a two phase selection. Allowing simple random sampling without replacement (SRSWOR) scheme at each phase, the two (or double) sampling scheme is as follows:

- (i) A first phase sample  $s'$  ( $s' \subset U$ ) of fixed size  $n$  is drawn from  $U$  to observe only  $x_1$  in order to find an estimate of  $\bar{X}_1$ .
- (ii) Given  $s'$ , a second phase sample  $s$  ( $s \subset s'$ ) of fixed size  $n$  is drawn from  $s'$  to observe  $y$  only.

Let  $\bar{X}_{1(2)} = \sum_{i \in s} x_{1i}/n$ ,  $\bar{Y}_{(2)} = \sum_{i \in s} y_i/n$  and  $\bar{X}_{1(1)} = \sum_{i \in s'} x_i/n$ . The two phase sampling ratio and regression estimators for  $\bar{Y}$  in this case, are defined by

$$\bar{y}_{Rd}^{(1)} = \frac{\bar{Y}_{(2)}}{\bar{X}_{1(2)}} \bar{X}_{1(1)} \tag{1.1}$$

and

$$\bar{y}_{ld}^{(1)} = \bar{Y}_{(2)} + \hat{\beta}_{yx_1}^{(2)} (\bar{X}_{1(1)} - \bar{X}_{1(2)}), \tag{1.2}$$

where  $\hat{\beta}_{yx_1}^{(2)} = \sum_{i \in s} (y_i - \bar{Y}_{(2)})(x_{1i} - \bar{X}_{1(2)}) / \sum_{i \in s} (x_{1i} - \bar{X}_{1(2)})^2$  is the estimate of the population regression coefficient  $\beta_{yx_1}$  of  $y$  on  $x_1$ . The approximate expressions for their mean square errors to terms of order  $n^{-1}$ , ignoring finite population correction (FPC) terms, are given by

$$\text{MSE}(\bar{y}_{Rd}^{(1)}) = \bar{Y}^2 \left[ \frac{C_y^2 - 2\rho_{yx_1} C_y C_{x_1} + C_{x_1}^2}{n} + \frac{2\rho_{yx_1} C_y C_{x_1} - C_{x_1}^2}{n'} \right] \tag{1.3}$$

and

$$\text{MSE}(\bar{y}_{ld}^{(1)}) = \bar{Y}^2 C_y^2 \left[ \frac{1 - \rho_{yx_1}^2}{n} + \frac{\rho_{yx_1}^2}{n'} \right], \tag{1.4}$$

where  $C_y = S_y/\bar{Y}$  and  $C_{x_1} = S_{x_1}/\bar{X}_1$  are the coefficients of variation of  $y$  and  $x_1$  respectively; and  $\rho_{yx_1} = S_{yx_1}/(S_y S_{x_1})$  is the correlation coefficient between  $y$  and  $x_1$ ,  $S_y^2 = \sum_{i \in U} (y_i - \bar{Y})^2 / (N - 1)$ ,  $S_{x_1}^2 = \sum_{i \in U} (x_{1i} - \bar{X}_1)^2 / (N - 1)$  and  $S_{yx_1} = \sum_{i \in U} (y_i - \bar{Y})(x_{1i} - \bar{X}_1) / (N - 1)$ .

In many practical situations even though  $\bar{X}_1$  is not known, information on  $\bar{X}_2$ , the population mean of another cheaply ascertainable variable  $x_2$ , closely related to  $x_1$  but compared to  $x_1$  remotely related  $y$ , is available. For example, while estimating the total yield of wheat in a district, the yield and area under the crop are likely to be unknown, but the total area of each farm may be known. Then  $y$ ,  $x_1$  and  $x_2$  are respectively the yield, area under wheat and area under cultivation, see Sahoo and Sahoo (1993). This type of situation has been also briefly discussed by Chand (1975), Kiregyera (1980, 1984), Srivenkataramana and Tracy (1989), Sahoo *et al.* (1993) and Sahoo and Sahoo (1993), among others. In this case the sampling model involving  $n'$  and  $n$  is as follows:

Select the first phase sample  $s'$  of size  $n'$  to observe  $x_1$  and  $x_2$  to furnish an estimate of  $\bar{X}_1$ . Then select a subsample  $s$  ( $s \subset s'$ ) of  $n$  units to observe  $y$  only. Chand (1975) suggested a chain ratio-in-ratio estimator of  $\bar{Y}$  as:

$$\bar{y}_{Rd}^{(2)} = \bar{Y}_{(2)} \left( \frac{\bar{X}_{1(1)}}{\bar{X}_{1(2)}} \right) \left( \frac{\bar{X}_2}{\bar{X}_{2(1)}} \right), \tag{1.5}$$

where  $\bar{X}_{2(1)} = \sum_{i \notin \epsilon s'} x_{2i} / n'$ .

The estimator  $\bar{y}_{Rd}^{(2)}$  has been further generalized by various authors including Kiregyera (1980), Srivastava *et al.* (1990), Singh and Singh (1991), Upadhyaya *et al.* (1992), Singh (1993), Singh *et al.* (1992), Sahoo *et al.* (1993) and others. When  $\bar{X}_2$ , the population mean of the second auxiliary variable  $x_2$  is known Kiregyera (1984) obtained a ratio-in-regression estimator

$$\bar{y}_{ld}^{(2)} = \bar{Y}_{(2)} + \hat{\beta}_{yx_1}^{(2)} \left( \frac{\bar{X}_{1(1)}}{\bar{X}_{2(1)}} \bar{X}_2 - \bar{X}_{2(2)} \right) \tag{1.6}$$

and a regression-in-regression estimator

$$\bar{y}_{ld}^{(3)} = \bar{Y}_{(2)} + \hat{\beta}_{yx_1}^{(2)} [(\bar{X}_{1(1)} - \bar{X}_{1(2)}) - \hat{\beta}_{x_1x_2}^{(1)} (\bar{X}_{2(1)} - \bar{X}_2)], \tag{1.7}$$

where  $\hat{\beta}_{x_1x_2}^{(1)} = \sum_{i \notin \epsilon s'} (x_{1i} - \bar{X}_{1(1)})(x_{2i} - \bar{X}_{2(1)}) / \sum_{i \notin \epsilon s'} (x_{2i} - \bar{X}_{2(1)})^2$  is the estimate of the population regression coefficient  $\beta_{x_1x_2}$  of  $x_1$  on  $x_2$ . These estimators  $(\bar{y}_{ld}^{(2)}, \bar{y}_{ld}^{(3)})$  have been further generalized by Singh and Singh (1991) and Upadhyaya *et al.* (1992). Assuming that the population mean  $\bar{X}_2$  and the variance  $S_{x_2}^2$  of the second auxiliary variable  $x_2$  are known, Tracy and Singh (1999) suggested a family of estimators which is further generalized by Singh and Ruiz Espejo (2000). Das and Tripathi (1980) advocated that in many situations of practical importance the population mean  $\bar{X}_2$  and the population variance  $S_{x_2}^2$  of the auxiliary variate  $x_2$  may not be known, but the knowledge of the coefficient of variation  $C_{x_2}$  of  $x_2$ , may be available. For discussion on this subject the reader is referred to Searls (1964), Sen (1978), Khan (1968) and Govindarajulu and Sahai (1972). The survey statisticians may utilize this information to obtaining estimators for  $\bar{X}_1$  better than the usual mean estimator  $\bar{X}_{1(2)}$ . Thus, motivated by Das and Tripathi (1980), one can define a family of estimators for  $\bar{X}_1$  as

$$\hat{\hat{X}}_{1t} = \frac{\bar{X}_{1(1)} - t_1(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)}{(\hat{C}_{x_2(1)}^2 - t_2(\hat{C}_{x_2(1)}^2 - C_{x_2}^2))^\alpha} (C_{x_2}^2)^\alpha, \tag{1.8}$$

where  $\hat{C}_{x_2(1)}^2 = S_{x_2(1)}^2 / \bar{X}_{2(1)}^2$ ,  $S_{x_2(1)}^2 = \sum_{i \in s'} (x_i - \bar{X}_{2(1)})^2 / (n' - 1)$  and  $C_{x_2(1)}^2 = S_{x_2}^2 / \bar{X}_2^2$  is the square of known coefficient of variation of  $x_2$ ,  $t_1$  and  $t_2$  are either constants or statistics such that their means  $E(t_1)$  and  $E(t_2)$  exist and  $\alpha$  is a suitably chosen constant. In this paper, making use of  $\hat{\hat{X}}_{1t}$  in (1.1) in place of  $\bar{X}_{1(1)}$  we have suggested a family of chain ratio-type estimators and its properties are discussed.

2. A FAMILY OF CHAIN RATIO-TYPE ESTIMATORS OF THE POPULATION MEAN

We define the following chain ratio-type estimators for the population mean  $\bar{Y}$  as

$$\bar{y}_{Rd}^{(t)} = \left( \frac{\bar{Y}_{(2)}}{\bar{X}_{1(2)}} \right) \frac{\bar{X}_{1(1)} - t_1(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)}{\hat{C}_{x_2}^2 - t_2(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)} (C_{x_2}^2)^\alpha, \tag{2.1}$$

where  $\alpha, t_1$  and  $t_2$  are the same as those defined in Section 1 for the estimator  $\bar{X}_{1t}$ . To study the properties of the family  $\bar{y}_{Rd}^{(t)}$  in (2.1), we impose the following restrictions on the choices of  $t_1$  and  $t_2$ :

- (i) Either  $E(t_1)$  and  $E(t_2)$  do not depend on sample sizes  $(n, n')$  or they are such that  $E(t_i) = E_0(t_i) + o(n^{-l}), l > 0, i = 1, 2$ , where  $E_0(t_i)$  is a constant (parameter) not depending on the sample sizes  $(n, n')$ .
- (ii) In the case that  $t_1$  and  $t_2$  are not constants they are such that  $V(t_i), (i = 1, 2)$ ; and  $\text{Cov}(t_1, t_2)$  are of order and the covariances between  $t_i$  and any of the  $\bar{Y}_{(2)}, \bar{X}_{1(1)}, \bar{X}_{1(2)}, \bar{X}_{2(1)}, \bar{X}_{2(2)}$  and  $S_{x(2)}^2$  and all the higher order moments are of order  $n^{-1}$ , where  $l > 1$ .

For simplicity we assume that the population size  $N$  is large enough as compared to the sample sizes  $n$  and  $n'$  so that the finite population correction terms  $n/N$  and  $n'/N$  are ignored. We write

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{200}}{\bar{Y}^2}, \quad C_{x_1}^2 = \frac{S_{x_1}^2}{\bar{X}_1^2} = \frac{\mu_{020}}{\bar{X}_1^2}, \quad C_{x_2}^2 = \frac{S_{x_2}^2}{\bar{X}_2^2} = \frac{\mu_{002}}{\bar{X}_2^2},$$

$$\rho_{yx_1} = \frac{\mu_{110}}{\mu_{200}\mu_{020}}, \quad \rho_{yx_2} = \frac{\mu_{101}}{\mu_{200}\mu_{002}}, \quad \rho_{x_1x_2} = \frac{\mu_{011}}{\mu_{020}\mu_{002}}, \quad \lambda_{pqr} = \frac{\mu_{pqr}}{\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}}$$

and

$$\mu_{pqr} = \frac{1}{N} \sum_{i \in U} (y_i - \bar{Y})^p (x_{1i} - \bar{X}_1)^q (x_{2i} - \bar{X}_2)^r,$$

where  $(p, q, r)$  are non-negative integers. Further, we write

$$\bar{Y}_{0(2)} = \bar{Y}(1 + e_{1(1)}), \quad \bar{X}_{1(1)} = \bar{X}_1(1 + e_{0(2)}), \quad \bar{X}_{1(2)} = \bar{X}_1(1 + e_{1(2)}),$$

$$\bar{X}_{2(1)} = \bar{X}_2(1 + e_{2(1)}), \quad \bar{X}_{2(2)} = \bar{X}_2(1 + e_{2(2)}), \quad S_{x_2(1)}^2 = S_{x_2}^2(1 + e_{3(1)}),$$

where

$$\begin{aligned}
 E(e_{0(2)}) &= E(e_{1(1)}) = E(e_{1(2)}) = E(e_{2(1)}) = E(e_{2(2)}) = E(e_{3(1)}) = 0, \\
 E(e_{0(2)}^2) &= \frac{1}{n}C_y^2, \quad E(e_{1(1)}^2) = \frac{1}{n'}C_{x_1}^2, \quad E(e_{1(2)}^2) = \frac{1}{n}C_{x_1}^2, \quad E(e_{2(1)}^2) = \frac{1}{n'}C_{x_2}^2, \\
 E(e_{2(2)}^2) &= \frac{1}{n}C_{x_2}^2, \quad E(e_{0(2)}e_{1(1)}) = \frac{1}{n'}\rho_{yx_1}C_{x_1}C_y, \quad E(e_{0(2)}e_{1(2)}) = \frac{1}{n}\rho_{yx_1}C_yC_{x_1}, \\
 E(e_{0(2)}e_{2(1)}) &= \frac{1}{n'}\rho_{yx_2}C_yC_{x_2}, \quad E(e_{0(2)}e_{2(2)}) = \frac{1}{n}\rho_{yx_2}C_yC_{x_2}, \\
 E(e_{1(1)}e_{2(2)}) &= \frac{1}{n'}\rho_{x_1x_2}C_{x_1}C_{x_2}, \quad E(e_{1(2)}e_{2(1)}) = \frac{1}{n'}\rho_{x_1x_2}C_{x_1}C_{x_2}, \\
 E(e_{1(2)}e_{2(2)}) &= \frac{1}{n}\rho_{x_1x_2}C_{x_1}C_{x_2}, \quad E(e_{2(1)}e_{2(2)}) = \frac{1}{n'}C_{x_2}^2
 \end{aligned}$$

and to the first-degree of approximation

$$\begin{aligned}
 E(e_{0(2)}e_{3(1)}) &= \frac{1}{n'}\lambda_{102}C_y, \quad E(e_{1(1)}e_{3(1)}) = \frac{1}{n'}\lambda_{012}C_{x_1}, \\
 E(e_{1(2)}e_{3(1)}) &= \frac{1}{n'}\lambda_{012}C_{x_1}, \quad E(e_{2(1)}e_{3(1)}) = \frac{1}{n'}\lambda_{003}C_{x_2}, \\
 E(e_{2(2)}e_{3(1)}) &= \frac{1}{n'}\lambda_{003}C_{x_2}, \quad E(e_{3(1)}^2) = \frac{1}{n'}(\lambda_{004} - 1).
 \end{aligned}$$

Assuming  $|(\bar{X}_{1(2)} - \bar{X}_1)/\bar{X}_1| < 1$ ,  $|(\bar{X}_{2(1)} - \bar{X}_2)/\bar{X}_2| < 1$  and  $|(1 - t_2)(\hat{C}_{x_{2(1)}}^2 - C_{x_2}^2)/C_{x_2}^2| < 1$ , the bias  $B(\bar{y}_{Rd}^{(t)})$  and mean square error (MSE) of the estimator  $\bar{y}_{Rd}^{(t)}$  in (2.1), to terms of order  $o(n^{-1})$ , are given by

$$\begin{aligned}
 B(\bar{y}_{Rd}^{(t)}) &= \left(\frac{1}{n} - \frac{1}{n'}\right)\bar{Y}(C_{x_1}^2 - \rho_{yx_1}C_{x_1}C_y) + \left(\frac{\bar{Y}}{n'}\right)\left[\left\{\frac{\alpha(\alpha + 1)}{2}(1 - E_o(t_2))^2\right.\right. \\
 &\quad \left.\left. + \left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o(t_2)\alpha(1 - E_o(t_2))\right\}(4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1)\right. \\
 &\quad \left. - \left\{\alpha(1 - E_o(t_2)) + \left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o(t_1)\right\}(3C_{x_2}^2 - 2\lambda_{003}C_{x_2})\right. \\
 &\quad \left. - \alpha(1 - E_o(t_2))C_y(\lambda_{102} - 2\rho_{yx_2}C_{x_2}) - \frac{E_o(t_1)C_{x_2}^2}{\bar{X}_1}\{C_y(\lambda_{102} - 2\rho_{yx_2}C_{x_2})\right. \\
 &\quad \left. - C_{x_1}(\lambda_{102} - 2\rho_{x_1x_2}C_{x_2})\right] + \bar{Y}\alpha E_o(t_2)\{C(t_2, S_{x_{2(1)}}^2) - 2C(t_2, \bar{X}_{2(1)})\} \\
 &\quad - \left(\frac{C_{x_2}^2}{\bar{X}_1}\right)\bar{Y}E_o(t_1)\{C(t_1, S_{x_2}^2) - 2C(t_1, \bar{X}_{2(1)})\} \tag{2.2}
 \end{aligned}$$

and

$$\begin{aligned} \text{MSE}(\bar{y}_{Rd}^{(t)}) &= \text{MSE}(\bar{y}_{Rd}^{(1)}) + \left(\frac{\bar{Y}^2}{n'}\right) \left\{ \left\{ \left(\frac{C_{x_2}^2}{\bar{X}_1}\right) E_o(t_2) \right. \right. \\ &\quad \left. \left. + \alpha(1 - E_o(t_2)) \right\}^2 (4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1) \right. \\ &\quad \left. - 2\left\{ \left(\frac{C_{x_2}^2}{\bar{X}_1}\right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right\} \right. \\ &\quad \left. C_y(\lambda_{012} - 2\rho_{yx_2}C_{x_2}) \right\}, \end{aligned} \tag{2.3}$$

where

$$\begin{aligned} C(t_i, \bar{X}_{2(1)}) &= \frac{\text{Cov}(t_i, \bar{X}_{2(1)})}{\bar{X}_2 E_o(t_i)}, \quad i = 1, 2, \\ C(t_i, S_{x_2(1)}^2) &= \frac{\text{Cov}(t_i, S_{x_2(1)}^2)}{S_{x_2}^2 E_o(t_i)}, \quad i = 1, 2 \end{aligned}$$

and  $\text{MSE}(\bar{y}_{Rd}^{(1)})$  is given by (1.3). The optimum values of  $E_o(t_1)$ ,  $E_o(t_2)$  and  $\alpha$  which minimize  $\text{MSE}(\bar{y}_{Rd}^{(t)})$  are given by

$$\left\{ \left(\frac{C_{x_2}^2}{\bar{X}_1}\right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right\} = \frac{C_y(\lambda_{012} - 2\rho_{yx_2}C_{x_2})}{(4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1)} = \alpha_0 \tag{2.4}$$

and thus the resulting minimum mean square error of  $\bar{y}_{Rd}^{(t)}$  is given by

$$\min \text{MSE}(\bar{y}_{Rd}^{(t)}) = \text{MSE}(\bar{y}_{Rd}^{(1)}) - \left(\frac{S_y^2}{n'}\right) \frac{(\lambda_{012} - 2\rho_{yx_2}C_{x_2})^2}{(4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1)}, \tag{2.5}$$

where  $\text{MSE}(\bar{y}_{Rd}^{(1)})$  is given in (1.3).

REMARK 2.1.

- (i) One may generate a large number of estimators from (2.1) by putting suitable choices of  $t_1$ ,  $t_2$  and  $\alpha$ .
- (ii) In addition to the family of estimators

$$\bar{y}_{Rd}^{(3)} = \bar{Y}_{(2)} \frac{\{\bar{X}_{1(1)} - t_1(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)\}}{\bar{X}_{1(2)}},$$

many other estimators may be identified as a subclass of the family in (2.1). The bias and MSE of  $\bar{y}_{Rd}^{(3)}$  can be obtained from (2.2) and (2.3) respectively just putting  $\alpha = 0$ .

- (iii) The term  $\alpha_0 = C_y(\lambda_{102} - 2\rho_{yx_2} C_{x_2}) / (4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1)$  given in (2.4) may be written as

$$\alpha_0 = \frac{C_y \{ \rho^* \sqrt{(\lambda_{004} - 1)} - \rho_{yx_2} C_{x_2} \}}{4C_{x_2}(C_{x_2} - \lambda_{003}) + \lambda_{004} - 1}, \tag{2.6}$$

where  $\rho^*$  is the correlation coefficient between  $y$  and  $(x_2 - x_1)^2$ .

- (iv) The estimators  $\bar{y}_{Rd}^{(t)}$  in (2.1) satisfying (2.4) will be termed as the estimators belonging to the asymptotically optimal subclass of  $\bar{y}_{Rd}^{(t)}$ .

### 3. EFFICIENCY COMPARISON

From (2.3) we find that the estimator  $\bar{y}_{Rd}^{(t)}$  will be better than the usual ratio estimator in two phase sampling if  $t_1, t_2$  and  $\alpha$  are chosen in such a manner that

$$\left[ \left( \frac{C_{x_2}^2}{\bar{X}_1} \right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right] \tag{3.1}$$

lies between 0 and  $2\alpha_0$ . It is well known that under the SRSWOR sampling scheme (ignoring FPC)

$$V(\bar{Y}_{(2)}) = \frac{1}{n} \bar{Y}^2 C_y^2. \tag{3.2}$$

It follows from (1.3) and (3.2) that the usual ratio estimator in two phase sampling is more efficient than the conventional unbiased estimator  $\bar{Y}_{(2)}$  if

$$\rho_{yx_1} > \frac{1}{2} \frac{C_{x_1}}{C_y}. \tag{3.3}$$

Thus combining (3.1) and (3.3) we establish that the proposed family of estimators  $\bar{y}_{Rd}^{(t)}$  would be more efficient than  $\bar{Y}_{(2)}$  if and only if  $t_1, t_2$  and  $\alpha$  are chosen in such a way that

$$\left[ \left( \frac{C_{x_2}^2}{\bar{X}_1} \right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right] \tag{3.4}$$

lies between 0 and  $2\alpha_0$  with  $\rho_{yx_1} > C_{x_1}/2C_y$ .

REMARK 3.1. Let  $\alpha$  and  $t_2 = \alpha_1$  (a constant) be preassigned. Then for  $t_1 = w\bar{X}_{1(1)}/\hat{C}_{x_2}^2$  and  $t_1 = -w\bar{X}_{1(1)}/\hat{C}_{x_2}^2$  the proposed family  $\bar{y}_{Rd}^{(t)}$  yields the subclasses

$$y_{Rd}^{(4)} = \left( \frac{\bar{Y}_{(2)}}{\bar{X}_{1(2)}} \right) \frac{(1-w)\bar{X}_{1(1)} + w\bar{X}_{1(1)} \frac{C_{x_2}^2}{\hat{C}_{x_2}^2}}{\{(1-\alpha_1)\hat{C}_{x_2(1)}^2 + \alpha_1 C_{x_2}^2\}^\alpha} \left( C_{x_2}^2 \right)^\alpha \tag{3.5}$$



and

$$\bar{y}_{Rd}^{(5)} = \left( \frac{\bar{Y}_{(2)}}{\bar{X}_{1(2)}} \right) \frac{(1+w)\bar{X}_{1(1)} - w\bar{X}_{1(1)} \frac{C_{x_2}^2}{\hat{C}_{x_2(1)}}}{\{(1-\alpha)\hat{C}_{x_2(1)}^2 + \alpha C_{x_2}^2\}^{\alpha_1}} \left( C_{x_2}^2 \right)^\alpha \tag{3.6}$$

respectively where  $w$  is a constant.

From (3.1) and (3.4) it follows that the estimators  $\bar{y}_{Rd}^{(4)}$  and  $\bar{y}_{Rd}^{(5)}$  are more efficient than the estimators  $\bar{y}_{Rd}^{(1)} = \bar{X}_{1(1)}\bar{Y}_{(2)}/\bar{X}_{1(2)}$  and  $\bar{Y}_{(2)}$  if and only if  $w$  lies between

$$-\alpha(1-\alpha) \text{ and } \{2\alpha_0 - \alpha(1-\alpha_1)\} \tag{3.7}$$

and

$$-\alpha(1-\alpha_1) \text{ and } \{2\alpha_0 - \alpha(1-\alpha_1)\} \text{ with } \rho_{yx_1} > \frac{1}{2} \frac{C_{x_1}}{C_y} \tag{3.8}$$

where  $\alpha_0$  is defined in (2.6).

REMARK 3.2. We consider an estimator of  $\bar{Y}$  as

$$\bar{y}_{Rd}^{(6)} = \bar{Y}_{(2)} \left( \frac{\bar{X}_{1(1)}}{\bar{X}_{1(2)}} \right) \frac{C_{x_2}^2}{\{(1-w_1)C_{x_2}^2 + w_1\hat{C}_{x_2(1)}^2\}}$$

which is also identified as a particular member of the family  $\bar{y}_{Rd}^{(t)}$ , where  $w_1$  is suitably chosen scalar. The MSE of  $\bar{y}_{Rd}^{(6)}$  to the first degree of approximation is given by

$$\begin{aligned} \text{MSE} \left( \bar{y}_{Rd}^{(6)} \right) &= \text{MSE} \left( \bar{y}_{Rd}^{(1)} \right) + \left( \frac{\bar{Y}^2}{n'} \right) [w_1^2(4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1) \\ &\quad - 2w_1(\lambda_{102} - 2\rho_{yx_2}C_{x_2})C_y], \end{aligned} \tag{3.9}$$

where  $\text{MSE} \left( \bar{y}_{Rd}^{(1)} \right)$  is given in (1.3). It follows from (3.9) that the estimator  $\bar{y}_{Rd}^{(6)}$  is better than the conventional ratio estimator  $\bar{y}_{Rd}^{(1)}$  in two phase sampling if either

$$0 < w_1 < 2\alpha_0 \text{ or } 2\alpha_0 < w_1 < 0. \tag{3.10}$$

In practice it would be difficult to get the exact optimum value  $\alpha_0$  in (2.4) as it involve unknown parameters  $C_y, C_{x_2}, \rho_{yx_2}, \lambda_{102}, \lambda_{003}$  and  $\lambda_{004}$ . However, one may have the good guessed approximate values of three parameters from a most recent survey taken in the past or by conducting a preliminary survey taken in the past or by conducting a preliminary survey by utilizing a small fraction of the budget allocated for the present survey or past data or experience or even from expert guesses by the specialists in the concerned field. Thus it is not entirely

unrealistic to assume prior knowledge of  $C_y$ ,  $C_{x_2}$ ,  $\rho_{yx_2}$ ,  $\lambda_{102}$ ,  $\lambda_{003}$  and  $\lambda_{004}$  and the value of  $\alpha_0$ . It follows from (3.10) that if one has only the good guessed approximate values of the parameters involved, the resulting estimator will be better than usual two phase sampling ratio estimator  $\bar{y}_{Rd}^{(1)}$ , for instance, see Das and Tripathi (1978, 1980) and Srivastara and Jhaji (1980, 1981). The estimator  $\bar{y}_{Rd}^{(6)}$  is better than the sample mean  $\bar{Y}_{(2)}$  if the condition (3.10) along with the condition  $\rho_{yx_2} = C_{x_2}/2C_y$  hold.

REMARK 3.3. Let us consider the distribution in which  $\mu_{102} = 0$  and  $\mu_{003} = 0$ . Let  $C_y$  and  $C_{x_2}$  be positive and  $\bar{X}_1^{(1)}$ ,  $\rho_{yx_2}^{(1)}$ ,  $\rho_{yx_2}^{(2)}$ ,  $C_y^{(1)}$  and  $\lambda_{004}^{(1)}$  be quantities such that  $0 < \bar{X}_1^{(1)} \leq \bar{X}_1$ ,  $0 < \rho_{yx_2}^{(1)} < \rho_{yx_2}$ ,  $\rho_{yx_2} < \rho_{yx_2}^{(2)} < 0$ ,  $0 < C_y^{(1)} < C_y$  and  $\lambda_{004} \leq \lambda_{004}^{(1)}$  (Das and Tripathi, 1980). With the help of (3.10) a set of sufficient condition for  $\bar{y}_{Rd}^{(6)}$  to be better than  $\bar{y}_{Rd}^{(1)}$  is given by

$$0 < w_1 < 2\alpha_0^* \tag{3.11}$$

in case  $\rho_{yx_2}$  is negative, and

$$2\alpha_0^{**} < w_1 < 0 \tag{3.12}$$

in case  $\rho_{yx_2}$  is positive, where

$$\alpha_1^{**} = -\frac{2\rho_{yx_2}^{(2)}C_y^{(1)}C_{x_2}}{(4C_{x_2}^2 + \lambda_{004}^{(1)} - 1)} \tag{3.13}$$

and

$$\alpha_0^{**} = -\frac{2\rho_{yx_2}^{(1)}C_y^{(1)}C_{x_2}}{(4C_{x_2}^2 + \lambda_{004}^{(1)} - 1)}. \tag{3.14}$$

REMARK 3.4. In the case of the trivariate normal distribution, the MSE of the proposed family  $\bar{y}_{Rd}^{(t)}$  in (2.3) reduces to

$$\begin{aligned} \text{MSE}\left(\bar{y}_{Rd}^{(t)}\right) &= \text{MSE}\left(\bar{y}_{Rd}^{(1)}\right) \\ &+ 2\left(\frac{\bar{Y}^2}{n'}\right)\left[\left(1 + C_{x_2}^2\right)\left\{\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o(t_1) + \alpha(1 - E_o(t_2))\right\}^2\right. \\ &\left.+ 2\rho_{yx_2}C_yC_{x_2}\left\{\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o(t_1) + \alpha(1 - E_o(t_2))\right\}\right] \end{aligned} \tag{3.15}$$

which is minimized for

$$\left\{\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o(t_1) + \alpha(1 - E_o(t_2))\right\} = -\frac{2\rho_{yx_2}C_yC_{x_2}}{(1 + 2C_{x_2}^2)}. \tag{3.16}$$

Thus the resulting minimum MSE is given by

$$\min \text{MSE}(\bar{y}_{Rd}^{(t)}) = \text{MSE}(\bar{y}_{Rd}^{(1)}) - \frac{2S_y^2 \rho_{yx_2}^2 C_{x_2}^2}{(1 + 2C_{x_2}^2)} \tag{3.17}$$

which clearly indicates that the proposed family  $\bar{y}_{Rd}^{(t)}$  has less minimum MSE than the MSE of  $\bar{y}_{Rd}^{(1)}$  in the trivariate normal distribution. It may be noted that the minimum MSE (3.17) is attained only when the optimum values of the scalars involved in the estimator  $\bar{y}_{Rd}^{(t)}$ , which are functions of unknown population parameters  $C_y$  and  $\rho_{yx_2}$  and known  $C_{x_2}$ , see (3.16), are used. It has been advocated by several authors that the value of  $C_y$  and  $\rho_{yx_2}$  are quite accurately known in practice, for instance, see Searls (1964), Khan (1968), Goundarajulu and Sahai (1972), Gleser and Healy (1976), Sen (1978), Lee (1981), Upadyyaya and Singh (1984), Srivastava (1967, 1971), Sahai and Sahai (1985), Singh (1986), Murthy (1967, pp. 96–99), Prasad (1989) and Reddy (1974, 1978). Thus the optimum value of scalars in (3.16) can be quite accurately known and hence the estimator  $\bar{y}_{Rd}^{(t)}$  will be attained the minimum MSE (3.17). It follows from (3.16) and above discussion that in the case of the trivariate normal distribution the knowledge of  $\rho_{yx_2}$ ,  $C_y$  and  $C_{x_2}$  is sufficient for the proposed estimator  $\bar{y}_{Rd}^{(t)}$  to be better than the usual two phase sampling ratio estimator  $\bar{y}_{Rd}^{(1)}$ . From (3.15) we note that

$$\begin{aligned} &\text{MSE}(\bar{y}_{Rd}^{(t)}) < \text{MSE}(\bar{y}_{Rd}^{(1)}) \\ \text{if } 0 < &\left\{ \left( \frac{C_{x_2}^2}{\bar{X}_1} \right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right\} \leq -\frac{2\rho_{yx_2} C_y C_{x_2}}{(1 + 2C_{x_2}^2)} \end{aligned} \tag{3.18}$$

in case  $\rho_{yx_2}$  is negative, and

$$-\frac{2\rho_{yx_2} C_y C_{x_2}}{(1 + 2C_{x_2}^2)} < \left\{ \left( \frac{C_{x_2}^2}{\bar{X}_1} \right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right\} < 0 \tag{3.19}$$

in case  $\rho_{yx_2}$  is positive.

#### 4. COMPARISON WITH SINGLE PHASE SAMPLING

Suppose that the samples are selected by SRSWOR. Then the appropriate estimator based on single-phase sampling without using any auxiliary variable is the mean per unit estimator  $\bar{Y}_{(2)}^*$ , where variance (ignoring FPC) is given by

$$V(\bar{y}_{Rd}^{(t)}) = \frac{1}{n^*} \bar{Y}^2 C_y^2, \tag{4.1}$$

where  $n^*$  is the sample size when no auxiliary variable is used. The comparison between two phase and single phase sampling can be made for fixed cost. For this purpose, let us consider the following cost function

$$C = C_1n + (C_2 + C_3)n', \tag{4.2}$$

where  $C_1, C_2$  and  $C_3$  are the cost per unit of collecting information on  $y, x_1$  and  $x_2$  respectively and  $C$  is the total cost of the survey.

The MSE of  $\bar{y}_{Rd}^{(t)}$  in (2.3) can be expressed as

$$\text{MSE}\left(\bar{y}_{Rd}^{(t)}\right) = \frac{M_1}{n} + \frac{M_2}{n'}, \tag{4.3}$$

where

$$\begin{aligned} M_1 &= \bar{Y}^2(C_y^2 + C_{x_1}^2 - 2\rho_{yx_1}C_yC_{x_1}), \\ M_2 &= \bar{Y}^2(2\rho_{yx_1}C_yC_{x_1} - C_{x_1}^2) + \gamma^2(4C_{x_2}^2 - 4\lambda_{003}C_{x_2} + \lambda_{004} - 1) \\ &\quad - 2\gamma C_y(\lambda_{102} - 2\rho_{yx_2}C_{x_2}) \end{aligned}$$

and

$$\gamma = \left[ \left( \frac{C_{x_2}^2}{\bar{X}_1} \right) E_o(t_1) + \alpha(1 - E_o(t_2)) \right].$$

The optimum values of  $n$  and  $n'$  for fixed cost  $C = C_0$  which minimizes the MSE of  $\bar{y}_{Rd}^{(t)}$  in (4.3) are given by

$$n_{opt} = \frac{C_0\sqrt{M_1/C_1}}{\sqrt{M_1C_1} + \sqrt{(C_2 + C_3)M_2}} \quad \text{and} \quad n'_{opt} = \frac{C_0\sqrt{M_2/(C_2 + C_3)}}{\sqrt{M_1C_1} + \sqrt{(C_2 + C_3)M_2}}. \tag{4.4}$$

Thus the resulting MSE of  $\bar{y}_{Rd}^{(t)}$  is given by

$$\text{MSE}_{opt}\left(\bar{y}_{Rd}^{(t)}\right) = \frac{1}{C_0} \left[ \sqrt{M_1C_1} + \sqrt{(C_2 + C_3)M_2} \right]^2. \tag{4.5}$$

If all resources are devoted instead to single phase sampling with no auxiliary variables, then the size of the sample is  $n^* = C_0/C_1$ , with variance of the sample mean as

$$V_{opt}\left(\bar{Y}_{(2)}^*\right) = \frac{C_1}{C_0} \bar{Y}^2 C_y^2. \tag{4.6}$$

Thus the proposed two-phase sampling strategy would be beneficial so long as

$$\text{MSE}_{opt}\left(\bar{y}_{Rd}^{(t)}\right) < V_{opt}\left(\bar{Y}_{(2)}^*\right),$$

that is

$$\frac{C_2 + C_3}{C_1} < \left[ \frac{S_y - \sqrt{M_1}}{\sqrt{M_2}} \right]^2 \tag{4.7}$$

Further the MSE of  $\bar{y}_{Rd}^{(1)}$  can be written as

$$\text{MSE}(\bar{y}_{Rd}^{(t)}) = \frac{M_1}{n} + \frac{V_0 - M_1}{n'} \tag{4.8}$$

where  $V_0 = \bar{Y}^2 C_y^2$ .

The optimum values of  $n$  and  $n'$  for fixed cost  $C_0 = nC_1 + n'C_2$  which minimizes the MSE of  $\bar{y}_{Rd}^{(1)}$  in (4.8)

$$n_{opt} = \frac{C_0 \sqrt{M_1/C_1}}{\sqrt{M_1 C_1} + \sqrt{(V_0 - M_1)C_2}} \text{ and } n'_{opt} = \frac{C_0 \sqrt{(V_0 - M_1)/C_2}}{\sqrt{M_1 C_1} + \sqrt{(V_0 - M_1)C_2}} \tag{4.9}$$

Thus the resulting MSE of  $\bar{y}_{Rd}^{(1)}$  is given by

$$\text{MSE}_{opt}(\bar{y}_{Rd}^{(1)}) = \frac{1}{C_0} \left[ \sqrt{M_1 C_1} + \sqrt{(V_0 - M_1)C_2} \right]^2 \tag{4.10}$$

From (4.5) and (4.10) we have that

$$\text{MSE}_{opt}(\bar{y}_{Rd}^{(t)}) < \text{MSE}_{opt}(\bar{y}_{Rd}^{(1)})$$

if

$$\frac{C_2 + C_3}{C_2} < \frac{V_0 - M_1}{M_2} \tag{4.11}$$

In particular, if we use the optimum estimator  $\bar{y}_{Rd}^{(t_0)}$  then the conditions in (4.7) and (4.11) for  $\bar{y}_{rd}^{(t_0)}$  to be better than  $\bar{Y}_{(2)}^*$  and  $\bar{y}_{Rd}^{(1)}$  respectively reduce to,

$$\frac{C_2 + C_3}{C_1} < \left[ \frac{S_y - \sqrt{M_1}}{\sqrt{M_2^*}} \right]^2 \tag{4.12}$$

and

$$\frac{C_2 + C_3}{C_2} < \left[ \frac{S_y^2 - M_1}{M_2^*} \right] \tag{4.13}$$

where

$$M_2^* = \left[ \bar{Y}^2 (2\rho_{yx_1} C_y C_{x_1} - C_{x_1}^2) - \frac{S_y^2 (\lambda_{102} - 2\rho_{yx_2} C_{x_2})^2}{(4C_{x_2} - 4\lambda_{003} C_{x_2} + \lambda_{004} - 1)} \right]$$

In case of trivariate normal distribution the inequalities in (4.12) and (4.13) respectively simplify to

$$\frac{C_2 + C_3}{C_1} < \left[ \frac{S_y - \sqrt{M_1}}{\sqrt{M_2^{**}}} \right]^2 \quad (4.14)$$

and

$$\frac{C_2 + C_3}{C_2} < \frac{S_y^2 - M_1}{M_2^{**}}, \quad (4.15)$$

where

$$M_2^{**} = \left[ \bar{Y}^2 (2\rho_{yx_1} C_y C_{x_1} - C_{x_1}^2) - \frac{2\rho_{yx_2}^2 C_{x_2}^2 C_y^2}{(1 + 2C_{x_2}^2)} \right].$$

REMARK 4.1. Motivated by Srivastava (1971, 1980), one can define an alternative family of estimators to  $\bar{y}_{Rd}^{(t)}$  in (2.1) as

$$\bar{y}_{Rd}^{(a)} = \frac{\bar{Y}_{(2)}}{\bar{X}_{1(2)}} L(\bar{X}_{1(1)}, \theta), \quad (4.16)$$

where  $\theta = \hat{C}_{x_2(1)}^2 / C_{x_2}^2$  and  $L(\bar{X}_{1(1)}, \theta)$  is a function of  $\bar{X}_{1(1)}$  and  $\theta$  such that  $L(\bar{X}_1, 1) = \bar{X}_1$  implies  $L_1(\bar{X}_1, 1) = \delta L(\bar{X}_{1(1)}, u) / \delta \bar{X}_{1(1)}|_{(\bar{X}_1, 1)} = 1$  and satisfies certain regularity conditions similar to those given in Srivastava (1971, 1980). To the first degree of approximation, it can be shown that the minimum MSE of  $\bar{y}_{Rd}^{(a)}$  is the same as that of  $\bar{y}_{Rd}^{(t)}$ , i.e.,

$$\min \text{MSE}(\bar{y}_{Rd}^{(a)}) = \min \text{MSE}(\bar{y}_{Rd}^{(t)}), \quad (4.17)$$

where minimum MSE of  $\bar{y}_{Rd}^{(t)}$  is given in (2.5).

## 5. A MORE GENERAL CLASS OF ESTIMATORS

We define a more general class of estimators of  $\bar{Y}$  as

$$\bar{y}_{Rd}^{(g)} = \left( \frac{\{\bar{X}_{1(1)} - t_1(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)\}}{\bar{X}_{1(2)}} \right)^{\alpha_1} \left( \frac{C_{x_2}^2}{\{\hat{C}_{x_2(1)}^2 - t_1(\hat{C}_{x_2(1)}^2 - C_{x_2}^2)\}} \right)^{\alpha_2}, \quad (5.1)$$

where  $\alpha_1$  and  $\alpha_2$  are suitably chosen constants and other notations are same as defined earlier. For  $(\alpha_1, \alpha_2) = (1, \alpha)$ ,  $\bar{y}_{Rd}^{(g)}$  reduces to the class of estimators  $\bar{y}_{Rd}^{(t)}$

defined in (2.1). To the first degree of approximation, the bias and MSE (ignoring FPC), are respectively given by

$$\begin{aligned}
 B\left(\bar{y}_{Rd}^{(g)}\right) = & \bar{Y}\left[\alpha_1\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)\left(C_{x_1}^2-\rho_{yx_1}C_{x_1}C_y\right)-\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)\frac{E_o\left(t_1\right)}{n'}A_1\right.\right. \\
 & \left.+\frac{\left(\alpha_1-1\right)}{2}A_2\right\}+\frac{\alpha_2 a_0}{n'}\left\{\frac{\left(\alpha_2+1\right)}{2}a_0 b-A_3\right\}+\alpha_1 \alpha_2 a_0 b\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)\frac{E_o\left(t_1\right)}{n'} \\
 & \left.-\alpha_1\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o\left(t_1\right)d_1+\alpha_2 E_o\left(t_2\right)d_2\right] \tag{5.2}
 \end{aligned}$$

and

$$\begin{aligned}
 \text{MSE}\left(\bar{y}_{Rd}^{(g)}\right) = & \bar{Y}^2\left[\frac{C_y^2}{n}+\left(\frac{1}{n}-\frac{1}{n'}\right)\alpha_1\left(\alpha_1-2\frac{\rho_{yx_1}C_y}{C_{x_1}}\right)C_{x_1}^2\right. \\
 & \left.+\frac{1}{n'}\left\{\alpha_1\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o\left(t_1\right)+\alpha_2 a_0\right\}\left\{\left(\alpha_1\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o\left(t_1\right)\right.\right.\right. \\
 & \left.\left.\left.+\alpha_2 a_0\right)-2C_y\left(\lambda_{102}-2\rho_{yx_2}C_{x_2}\right)\right\}\right], \tag{5.3}
 \end{aligned}$$

where

$$\begin{aligned}
 A_1 = & \left[A_3-C_{x_1}\left(\lambda_{102}-2\rho_{x_1x_2}C_{x_2}\right)\right], \quad A_2 = \left[\left(\frac{1}{n}-\frac{1}{n'}\right)C_{x_1}^2+\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)^2\left(\frac{b}{n'}\right)E_o\left(t_1\right)^2\right], \\
 A_3 = & \left[C_{x_2}\left(3C_{x_2}-2\lambda_{003}\right)+C_y\left(\lambda_{102}-2\rho_{x_1x_2}C_{x_2}\right)\right], \\
 a_0 = & 1-E_o\left(t_2\right), \quad b = \left(4C_{x_2}^2-4\lambda_{003}C_{x_2}+\lambda_{004}-1\right), \\
 d_1 = & \left\{C\left(t_1, S_{x_2(1)}^2\right)-2C\left(t_1, \bar{X}_{2(1)}\right)\right\}, \quad d_2 = \left\{C\left(t_2, S_{x_2(1)}^2\right)-2C\left(t_2, \bar{X}_{2(1)}\right)\right\}.
 \end{aligned}$$

The MSE  $\left(\bar{y}_{Rd}^{(g)}\right)$  in (5.3) is minimized for

$$\alpha_1 = \frac{\rho_{yx_1}C_y}{C_{x_1}} = K_{yx_1}$$

and

$$\left[\alpha_1\left(\frac{C_{x_2}^2}{\bar{X}_1}\right)E_o\left(t_1\right)+\alpha_2 E_o\left(t_2\right)\right] = \frac{C_y\left(\lambda_{102}-2\rho_{yx_2}C_{x_2}\right)}{b}. \tag{5.4}$$

Thus the resulting minimum MSE of  $\bar{y}_{Rd}^{(g)}$  is given by

$$\min \text{MSE}\left(\bar{y}_{Rd}^{(g)}\right) = S_y^2\left[\frac{1}{n}\left(1-\rho_{yx_1}^2\right)+\frac{1}{n'}\rho_{yx_1}^2-\frac{\left(\lambda_{102}-2\rho_{yx_1}C_{x_2}\right)^2}{n'\left(4C_{x_2}^2-4\lambda_{003}C_{x_2}+\lambda_{004}-1\right)}\right] \tag{5.5}$$

From (1.3), (1.4), (2.5) and (5.5) we have

$$\text{MSE}(\bar{y}_{Rd}^{(1)}) - \text{MSE}(\bar{y}_{ld}^{(1)}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \bar{Y}^2 C_{x_1}^2 (1 - K_{yx_1})^2 \geq 0, \quad (5.6)$$

$$\text{MSE}(\bar{y}_{ld}^{(1)}) - \min \text{MSE}(\bar{y}_{Rd}^{(g)}) = \frac{S_y^2 (\lambda_{102} - 2\rho_{yx_2} C_{x_2})^2}{n' (4C_{x_2}^2 - 4\lambda_{003} C_{x_2} + \lambda_{004} - 1)}, \quad (5.7)$$

$$\min \text{MSE}(\bar{y}_{Rd}^{(t)}) - \min \text{MSE}(\bar{y}_{Rd}^{(g)}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \bar{Y}^2 C_{x_1}^2 (1 - K_{yx_1})^2 \geq 0. \quad (5.8)$$

Thus we have the following inequalities:

$$\min \text{MSE}(\bar{y}_{Rd}^{(g)}) \leq \text{MSE}(\bar{y}_{ld}^{(1)}) \leq \text{MSE}(\bar{y}_{Rd}^{(1)}) \quad (5.9)$$

and

$$\min \text{MSE}(\bar{y}_{Rd}^{(g)}) \leq \min \text{MSE}(\bar{y}_{Rd}^{(t)}) \leq \text{MSE}(\bar{y}_{Rd}^{(l)}). \quad (5.10)$$

It follows from (5.9) and (5.10) that the estimator  $\bar{y}_{Rd}^{(g)}$  is more efficient than  $\bar{y}_{Rd}^{(l)}$ ,  $\bar{y}_{ld}^{(l)}$  and  $\bar{y}_{Rd}^{(t)}$  at optimum conditions.

REMARK 5.1. Following Srivastava and Jhajj (1981) we define an alternative class of estimators of population mean  $\bar{Y}$  as

$$\bar{y}_{Rd}^{(a^*)} = \bar{Y}_{(2)} G(u, \theta), \quad (5.11)$$

where  $u = \bar{X}_{1(2)}/\bar{X}_{1(1)}$ ,  $\theta = \hat{C}_{x_2(1)}^2/C_{x_2}^2$  and  $G(u, \theta)$  is a function of  $u$  and  $\theta$  such that  $G(1, 1) = 1$  and satisfies certain regularity conditions similar to those given in Srivastava and Jhajj (1981). To the first degree of approximation, it can be shown that the minimum MSE of  $\bar{y}_{Rd}^{(a^*)}$  is the same as that of  $\bar{y}_{Rd}^{(g)}$ , *i.e.*,

$$\min \text{MSE}(\bar{y}_{Rd}^{(a^*)}) = \min \text{MSE}(\bar{y}_{Rd}^{(g)}), \quad (5.12)$$

where  $\min \text{MSE}(\bar{y}_{Rd}^{(g)})$  is given in (5.5).

## 6. EMPIRICAL STUDY

The following two populations are considered to illustrate the relative behavior of the constructed estimators of  $\bar{Y}$ . From Cochran (1977), let  $y$  be the number of paralytic polio cases in ‘pacebo’ group,  $x_1$  be the number of paralytic polio cases in not inoculated group and  $x_2$  be the number of ‘pacebo’ children. Then, for this population,  $C_y = 1.2228$ ,  $C_{x_1} = 1.1317$ ,  $C_{x_2} = 1.0128$ ,  $\rho_{yx_1} = 0.6898$ ,  $\rho_{yx_2} = 0.7088$ ,  $\lambda_{003} = 2.3854$ ,  $\lambda_{004} = 8.9332$ ,  $\lambda_{102} = 0.5797$ ,  $n = 10$  and  $n' = 20$ .



We have computed the percent relative efficiencies of  $\bar{y}_{Rd}^{(1)}$ ,  $\bar{y}_{ld}^{(l)}$  and  $\bar{y}_{Rd}^{(t)}$  (at optimum condition) with respect to  $\bar{Y}_{(2)}$  and displayed in Table 6.1. Table 6.1 exhibits that the proposed family of estimator  $\bar{y}_{Rd}^{(t_0)}$  is more efficient than  $\bar{Y}_{(2)}$ ,  $\bar{y}_{Rd}^{(1)}$  and  $\bar{y}_{ld}^{(1)}$  with considerable gain. Thus the proposed estimator is recommended for its use in practice.

TABLE 6.1 Percent Relative Efficiencies of Different Estimators of  $\bar{Y}$  with respect to  $\bar{Y}_{(2)}$

Estimators	PRE(., $\bar{Y}_{(2)}$ )
$\bar{Y}_{(2)}$	100.00
$\bar{y}_{Rd}^{(1)}$	126.59
$\bar{y}_{ld}^{(1)}$	131.22
$\bar{y}_{Rd}^{(t)}$	157.38

REFERENCES

CHAND, L. (1975). "Some ratio-type estimators based on two or more auxiliary variables", Ph.D. Dissertation, Iowa State University, Iowa.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley & Sons, New Delhi.

DAS, A. K. AND TRIPATHI, T.P. (1978). "Use of auxiliary information in estimating the finite population variance", *Sankhya*, **40**, 139-148.

DAS, A. K. AND TRIPATHI, T. P. (1980). "Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known", *Sankhya*, **42**, 76-86.

GILESTER, L. J. AND HEALY, J. D. (1976). "Estimating the mean of a normal distribution with known coefficient of variation", *Journal of the American Statistics Association*, **71**, 977-981.

GOVINDARAJULU, Z. AND SAHAI, H. (1972). "Estimation of the parameters of a normal distribution with known coefficient of variation", *Reports of Statistical Application Research. Union of Japanese Scientists and Engineers*, **91**, 85-98.

KHAN, R. A. (1968). "A note on estimating the mean of a normal distribution with known coefficient of variation", *Journal of the American Statistics Association*, **63**, 1039-1041.

KIREGYERA, B. (1980). "A chain ratio-type estimator in finite population in double sampling using two auxiliary variables", *Metrika*, **27**, 217-223.

KIREGYERA, B. (1984). "Regression type estimators using two auxiliary variables and the model of double sampling from finite populations", *Metrika*, **31**, 215-226.

LEE, K. H. (1981). "Estimation of variance of mean using known coefficient of variation", *Communications in Statistics. Theory and Methods*, **10**, 503-514.

MURTHY, M. N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

NEYMAN, J. (1938). "Contribution to the theory of sampling human populations", *Journal of the American Statistics Association*, **33**, 101-116.

PRABHU-AJGAONKAR, S. G. (1975). "The efficient use of supplementary information in double sampling procedures", *Sankhya*, **37**, 181-189.

- PRASAD, B. (1989). "Some improved ratio type estimators of population mean and ratio in finite population sample surveys", *Communications in Statistics. Theory and Methods*, **18**, 379-392.
- REDDY, V. N. (1974). "A transformed ratio method of estimation", *Sankhya*, **36**, 59-70.
- REDDY, V. N. (1978). "A study on the use of prior knowledge on certain population parameters in estimation", *Sankhya*, **40**, 29-34.
- SAHAI, A. AND SAHAI, A. (1985). "On efficient use of auxiliary information", *Journal of Statistical Planning and Inference*, **12**, 203-212.
- SAHOO, J. AND SAHOO, L. N. (1993). "A class of estimators in two-phase sampling using two auxiliary variables", *Journal of the Indian Society of Agricultural Statistics*, **31**, 107-114.
- SAHOO, L. N., SAHOO, J. AND MOHANTY, S. (1993). "A regression approach to estimation in two-phase sampling using two auxiliary variables", *Current Science*, **65**, 73-75.
- SEARLS, D. T. (1964). "The utilization of known coefficient of variation in the estimation procedure", *Journal of the American Statistics Association*, **59**, 1225-1226.
- SEN, A. R. (1978). "Estimation of the population mean when the coefficient of variation is known", *Communications in Statistics. Theory and Methods*, **7**, 657-672.
- SINGH, H. P. (1986). "A note on the estimation of variance of sample mean using the knowledge of coefficient of variation in normal population", *Communications in Statistics. Theory and Methods*, **15**, 3737-3746.
- SINGH, V. K. AND SINGH, G. N. (1991). "Chain type regression estimators with two auxiliary variables under double sampling scheme", *Metron*, **49**, 279-289.
- SINGH, H. P., SINGH, H. P. AND KUSHWAHA, K. S. (1992). "On chain ratio-to-regression type estimator in double sampling", *Assam Statistical Review*, **6**, 91-105.
- SINGH, H. P. (1993). "A chain ratio-cum-difference estimator using two auxiliary variates in double sampling", *Journal of Ravishankar University*, Ser. B, **6**, 79-83.
- SINGH, H. P. AND ESPEJO, M. R. (2000). "An improved class of chain regression estimators in two phase sampling", *Statistics and Decisions*, **18**, 205-218.
- SRIVASTAVA, S. K. (1967). "An estimator using auxiliary information in sample surveys", *Calcutta Statistical Association Bulletin*, **16**, 121-131.
- SRIVASTAVA, S. K. (1971). "A generalized estimator for the mean of a finite population using multi-auxiliary information estimator using auxiliary information in sample surveys", *Journal of the American Statistics Association*, **66**, 404-407.
- SRIVASTAVA, S. K. (1980). "A class of estimators using auxiliary information in sample surveys", *Canadian Journal of Statistics*, **8**, 253-254.
- SRIVASTAVA, S. K. AND JHAJJ, H. S. (1980). "A class of estimators using auxiliary information for estimating finite population variance", *Sankhya*, **42**, 87-96.
- SRIVASTAVA, S. K. AND JHAJJ, H. S. (1981). "A class of estimators of the population mean in survey sampling using auxiliary information", *Biometrika*, **68**, 341-343.
- SRIVASTAVA, S. R., KHARE, B. B. AND SRIVASTAVA, S. R. (1990). "A generalized chain ratio estimator for mean of finite population", *Journal of the Indian Society of Agricultural Statistics*, **42**, 108-117.
- SRIVENKATARAMANA, T. AND TRACY, D. S. (1989). "Two-phase sampling for selection with probability proportional to size in sample surveys", *Biometrika*, **76**, 818-821.
- SUKHATME, B. V. (1962). "Some ratio-type estimators in two-phase sampling", *Journal of the American Statistics Association*, **57**, 628-632.

- TRACY, D. S. AND SINGH, H. P. (1999). "A general class of chain regression estimators in two phase sampling", *Journal of Applied Statistical Science*, **8**, 205–216.
- UNNIKRISHNAN, N. K. AND KUNTE, S. (1995). "Optimality of an analogue of Basu's estimator under a double sampling design", *Sankhya*, **57**, (1), 103–111.
- UPADHYAYA, L. N. AND SINGH, H. P. (1984). "On the estimator of the population mean with known coefficient of variation", *Biometrical Journal*, **26**, 8, 915–922.
- UPADHYAYA, L. N., DUBEY, S. P. AND SINGH, H. P. (1992). "A class of ratio in regression estimators using two auxiliary variables", *The Journal of Scientific Research*, **42**, 127–134.